Johnson Su-transformations for parameter estimation in ARMA-models when data are non-gaussian

van Montfort, R.; Schippers, R.; Heuts, R.M.J.

Publication date:
1983

subfaculteit der econometrie

RESEARCH MEMORANDUM

TILBURG UNIVERSITY
DEPARTMENT OF ECONOMICS
Postbus 90153 - 5000 LE Tilburg
Netherlands

JOHNSON S_U - TRANSFORMATIONS FOR PARAMETER ESTIMATION IN ARMA-MODELS WHEN DATA ARE NON-GAUSSIAN

R. van Montfort
R. Schippers
R. Heuts

JOHNSON S_U - TRANSFORMATIONS FOR PARAMETER ESTIMATION IN ARMA-MODELS WHEN DATA ARE NON-GAUSSIAN

R. VAN MONTFORT
R. SCHIPPERS
R. HEUTS

Abstract:

Correlated data are generated from ARMA time series models, varying the structure of the dependence among the observations, the kind of probability distributions, and the noise process variance. The inverse Johnson S_U transformation is used to generate non-Gaussian data.
Box-Jenkins parameter estimation is applied to (i) the original, non-Gaussian data, (ii) the back-transformed (normalized) data, and then a comparison is made between the quality of the parameter estimation results of these two approaches. It appears that the Box-Jenkins parameter estimation procedure is reasonably robust against non-normality of the Johnson S_U type. The power of two normality tests is examined for sample sizes of 50 and 200 correlated observations.
Further, there seems to be no clear relation between the quality of the parameter estimates of the Johnson S_U transformation and the quality of the structural parameter estimates in the ARMA model.
Keywords: Johnson S_U transforms, ARMA models, Monte Carlo experiments, tests of normality.

1. INTRODUCTION

In time series analysis an ARMA (p,q) process (see [1]) is represented by a stochastic difference equation of the type

$$\sum_{j=0}^{p} \phi_j x_{t-j} = \sum_{g=0}^{q} \theta_g \varepsilon_{t-g}, \quad \phi_0 = \theta_0 = 1 \tag{1}$$

For maximum likelihood estimation of the structural parameters in this equation it is usually assumed: $\varepsilon_t \in N(0, \sigma_\varepsilon^2)$. Under this assumption the process $\{x_t\}$ is also normally distributed: $x_t \in N(0,\gamma_0)$, where γ_0 is a function of $\phi_1,\ldots,\phi_p, \theta_1,\ldots,\theta_q, \sigma_\varepsilon^2$.

In practice (see e.g. [3], [5], [15]), however, process variables often deviate from normality. Empirical distributions usually are leptokurtic, which means that they have heavier tails than normal. The kurtosis will be large in those cases. In the present Monte Carlo study we examine to what extent the parameter estimation procedure of Box-Jenkins is insensitive to non-normality. Therefore we compare the following approaches: 1) parameter estimation using the original non-Gaussian data; 2) parameter estimation after transforming non-Gaussian into nearly Gaussian data via Johnson S_U transformations. For testing normality of correlated data we will use two statistics: Lomnicki's test statistic [14] and Cramèr-von Mises' test statistic (see e.g. [12]).

2. MONTE CARLO DESIGN

In this study we examined three different ARMA processes. The choice was determined by what we observed in practice for some series. The models are in operator form (where B is a backshift operator):

1) $(1 - 0.5B - 0.2B^2)(1 - 0.7B^{12})x_t = \varepsilon_t$

2) $1 + 0.7B + 0.4B^2 + 0.25B^3 + 0.12B^4)x_t = \varepsilon_t$

3) $x_t = (1 - 0.4B)(1 - 0.9B^{12})\ \varepsilon_t$

To generate sequences which not satisfy the normality assumption, we proceed as follows. We generate a sequence of 300 observations (the first 100 are discarded to eliminate the effect of the starting values $x_0,\dots,x_{-p+1}$) from each of the above schemes, where the noise process $\{\varepsilon_t\}$ is assumed $N(0,\sigma_\varepsilon^2)$. To generate normal variables we used the method of Brent [2]. The starting values $x_0,\dots,x_{-p+1}$ are in this case set to zero. We now have a sequence of x_t values with $x_t \in N(0,\gamma_0)$, where γ_0 is calculated for the three schemes in Table 1.

Table 1

Variances of $\{x_t\}$ processes for different schemes

Scheme	γ_0
1	$3.51268 * \sigma_\varepsilon^2$
2	$1.50579 * \sigma_\varepsilon^2$
3	$2.0996 * \sigma_\varepsilon^2$

In applying the Johnson S_U transformation (see [10]) we put a filter Φ on a non-normal process, from which a standard normal process originates:

$$\{x_t\} = \Phi\{y_t\},\ x_t \in N(0,1),\ y_t \in N,$$

where

$$\Phi\{y_t\} = \gamma + \delta \sinh^{-1} \{(y_t-\xi)/\lambda\} \tag{2}$$

and ξ, γ, λ and δ are the transformation parameters, which depend on the first four moments of y.

The inverse transformation is:

$$y_t = \Phi^{-1} \{x_t\} = \xi + \lambda \sinh \{(x_t - \gamma)/\delta\}, \tag{3}$$

and if we apply Φ^{-1}, the inverse Johnson S_U transformation, to a normal distributed process, we generate a non-normal process. In Appendix B another non-normal generation process is described, which seems, however, less appropriate in this case. We have taken three points in the transformation parameter space (ξ, γ, λ, δ).

The choice of these points in Table 2 is based on real-life data transformations by one of the authors [5].

Table 2

Selected transformation parameter values

parameter combination	ξ	γ	λ	δ
1	-0.005	-0.3	0.03	1.08
2	-0.001	-0.07	0.03	1.5
3	0	-1	1	0.5

In the above way we have generated non-normal sequences $\{y_t\} = \Phi^{-1} \{x_t\}$, with certain central moments μ_1^*, μ_2^*, μ_3^*, μ_4^*; these moments are functions of the parameters; see Appendix A. From the first four moments we can calculate a measure for skewness $\beta_1 = (\mu_3^*/(\mu_2^*)^{3/2})^2$ and a measure of kurtosis $\beta_2 = (\mu_4^*/(\mu_2^*)^2)$. Als β_1 and β_2 are dependent on the μ_i^*, and so on the transformation parameters and on γ_0 and σ_ε^2, we have to select an appropriate σ_ε^2. Table 3 must be read as follows: For the different values of σ_ε^2 in a row we can calculate the β_2-values for each scheme, and they are in the range as given in the table (the kurtosis of a normal distribution is 3).

Table 3

Choices for the noise variance σ_ε^2

β_2 \ parameter combination	1	2	3
low: 3.01-3.10	0.002	0.002	0.0005
medium: 4.70-12.4	0.25	0.4	0.025
high: 9.7-97	0.4	0.6	0.05

Summary:

- Generate for 3 different schemes and for 7 different variances σ_ε^2, 300 normally distributed x_t.
- Transform for 3 different schemes, 3 different transformation parameter combinations and 3 different variances, non-normally distributed data y_t.

3. TESTING FOR NORMALITY

As detection of non-normality is a first important step in data analysis, we now examine two test statistics for detecting non-normality of correlated data. Those test statistics will be applied to the y_t's for sample sizes of 50 and 200 observations.

Lomnicki [14] has proposed a test statistic which is based on the sample skewness and kurtosis and tests for departures from normality in the case of linear stochastic processes. He derived the asymptotic distribution of the sample skewness and kurtosis. The knowledge of these distributions allows us to test the departure from normality in the case of "large" samples, a problem which cannot be treated with the aid of classical tests based on the assumption that the sample values are independent. Let $\sqrt{b_1} = m_3/(m_2)^{3/2}$, $b_2 = m_4/(m_2^2) - 3$, where $m_r = \frac{1}{N} \sum_{t=1}^{N} (y_t - \bar{y})^r$, $r = 2,3,4$; and N the sample size. Lomnicki has shown that, for large N, if $\{y_t\}$ is Gaussian and stationary, $\sqrt{b_1}$ and b_2 are asymptotically normal, with: $E(\sqrt{b_1}) = E(b_2) = 0$, $\mathrm{var}(\sqrt{b_1}) = \frac{6}{N} \sum_{s=-\infty}^{\infty} \rho_y^3(s)$, $\mathrm{var}(b_2) = \frac{24}{N} \sum_{s=-\infty}^{\infty} \rho_y^4(s)$, where $\rho_y(s)$ is the autocorrelation coefficient of the process $\{y_t\}$ with lag s.

Since, in practice, $\rho_y(s)$ is unknown, the variances of $\sqrt{b_1}$ and b_2 are estimated by replacing $\rho_y(s)$ by $\hat{\rho}_y(s)$ (see e.g. [1]) and by replacing the infinite sum by a finite sum. So, $u_1 = (\sqrt{b_1})/\{\mathrm{var}(\sqrt{b_1})\}^{1/2}$, $u_2 = b_2/\{\mathrm{var}(b_2)\}^{1/2}$, are asymptotically distributed as standard normal variables.

The Cramèr-von Mises test statistic can also be used for testing normality when the data are correlated and the sample size is "large" (see e.g. Lawless [12]). One should be cautioned, however, that it is not known how large the sample size should be to make the classical quantiles reasonable for <u>correlated data</u>.

The test statistic is defined as

$$W_N^2 = \sum_{i=1}^{N} [\Phi((y_{(i)} - \hat{\mu})/\hat{\sigma}) - (i-0.5)/N] + \frac{1}{2N} ,$$

where N is the sample size, Φ is the distribution function of a standard normal variable, $y_{(i)}$ is the i-th ordered sample value, $\hat{\mu}$ is the sample mean, and $\hat{\sigma}$ is the sample standard deviation. In Table 4 the percentage points are listed for the W_N^2 test statistic and they are multiplied by a function of N, and so the percentage points are sufficiently accurate for virtually all N (see e.g. Stephens [16]):

Table 4

Percentage points for the modified W_N^2 test statistic

modified test statistic	percentage points				
$(1+0.5N^{-1})W_N^2$	0.75	0.90	0.95	0.975	0.99
	0.091	0.104	0.126	0.148	0.178

4. THE JOHNSON S_U TRANSFORM

Through the Johnson S_U transformation we can transform the y_t's to a standard normal variable x_t (see e.g. Hill [6]). First, however, we have to estimate the transformation parameters $\xi,\gamma,\lambda,\delta$. Johnson [10] has described the method of moments: the first four sample moments of y_t will lead to the estimates $\hat{\xi},\hat{\gamma},\hat{\lambda},\hat{\delta}$ using a numerical procedure (see [7], [8], [11]). The method of maximum likelihood is rather difficult to apply in this case.
We note that the fitted x_t is N(0,1) and the original x_t is N(0,γ_0), so that the transformation parameter estimates do not estimate the same parameters as in Table 2, but resp. $\gamma/\sqrt{\gamma_0}$, ξ, λ, $\delta/\sqrt{\gamma_0}$ (see APPENDIX A). An estimate for $(\gamma,\xi,\lambda,\delta)$ is then $(\sqrt{\gamma_0}\ \hat{\gamma},\ \hat{\xi},\ \hat{\lambda},\ \sqrt{\gamma_0}\ \hat{\delta})$.
After the calculation of these parameters one can calculate the fitted x_t as $\hat{\Phi}\{y_t\}$, where $\hat{\Phi}$ corresponds with the estimated parameters $(\hat{\gamma},\hat{\xi},\hat{\lambda},\hat{\delta})$. This is done for sample sizes of 50 and 200 observations.

5. BOX-JENKINS STRUCTURAL PARAMETER ESTIMATES AND THE NUMBER OF REPLICATIONS

With the computer program of Jenkins and Partners [9] we are able to calculate the approximate maximum likelihood estimates of the structural parameters of the different models of equation (1) given the model structure and the sample size of 50 resp. 200. We considered parameter estimates based on non-normal

y_t's and these estimates were compared with estimates based on Johnson S_U transformed data (the non-normal data are then transformed to standard normal data). We prefer that estimation procedure which leads to structural parameter estimates closest to the real parameter values.

This approach will give us some insight into the robustness of the parameter estimation procedure of Box and Jenkins. To get more reliable conclusions, we repeated the experiment a number of times. This number of replications was set to 10, because of the long computer time of the total experiment (approximately 12.000 sec. for batch plus 12.000 sec. for terminal use on an ICL 2960 computer). The estimated variances are then a measure of reliability.

6. NUMERICAL RESULTS

a. power results when testing normality

To get an impression of the quality of the two test statistics, we investigated the β-error in relation to the kurtosis, where β-error ≡ $P\{\text{accept } H_0 \mid H_1 \text{ is true}\}$, and H_0-hypothesis: normality (kurtosis = 3), H_1-hypothesis: non-normality (kurtosis ≠ 3). The β-error will be estimated by the fraction of acceptances of the H_0-hypothesis: $\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} b_i$, where n is the number of replications (10 in this case) and $b_i = 1$, when accepting H_0 and $b_i = 0$ when rejecting H_0.

Obviously $\hat{\beta}$ is an unbiased estimator of β. The level of the tests will be taken at 5%.

Table 5 shows the estimates of the β-errors for different cases listed for the test statistics of Lomnicki and Cramèr-von Mises, resp. for sample sizes of N = 50 and N = 200. For the noise variance we have three situations: L = low, M = medium, H = high. In order to analyze the results they are put together in

Figures 1 and 2. When comparing the two figures, we notice that the estimates of the β-errors are much lower for a sample size of 200 than one of 50, as we expected. And for a sample size of 200 the estimated β-errors are acceptable for both tests; however, for N = 50 even for high kurtosis values the β-errors are high.

Table 5

Estimated β-errors for the test statistics of Lomnicki and Cramèr-von Mises

SCHEME	TRANSFORMATION PARAMETER COMBIN.	VARIANCE	KURTOSIS	LOMNICKI $\hat{\beta}(N = 50)$	LOMNICKI $\hat{\beta}(N = 200)$	CRAMÈR-VON MISES $\hat{\beta}(N = 50)$	CRAMÈR-VON MISES $\hat{\beta}(N = 200)$	
1	1	L	3.03	1.0	1.0	1.0	0.7	
1	1	M	20.1	0.7	0.4	0.5	0.1	
1	1	H	96.7	0.2	0.1	0.1	0.0	
1	2	L	3.01	1.0	1.0	1.0	0.7	
1	2	M	11.1	0.7	0.2	0.4	0.3	
1	2	H	29.5	0.5	0.0	0.4	0.0	
1	3	L	3.10	1.0	1.0	0.9	1.0	
1	3	M	12.4	0.8	0.0	0.5	0.0	
1	3	H	41.1	0.3	0.0	0.1	0.0	
2	1	L	3.01	1.0	1.0	1.0	1.0	
2	1	M	5.81	0.9	0.4	0.8	0.1	
2	1	H	9.73	0.5	0.0	0.6	0.0	
2	2	L	3.01	1.0	1.0	1.0	1.0	

contuation of Table 5

2	2	M	4.68	0.8	0.1	1.0	0.3	
2	2	H	6.25	0.7	0.2	0.7	0.3	
2	3	L	3.04	1.0	1.0	1.0	0.9	
2	3	M	5.88	0.7	0.3	0.4	0.0	
2	3	H	10.4	0.5	0.0	0.1	0.0	
3	1	L	3.02	1.0	1.0	1.0	1.0	
3	1	M	8.08	0.9	0.2	0.6	0.2	
3	1	H	18.1	0.4	0.0	0.7	0.0	
3	2	L	3.01	1.0	1.0	1.0	1.0	
3	2	M	5.86	0.6	0.1	0.8	0.3	
3	2	H	9.31	0.6	0.2	1.0	0.3	
3	3	L	3.06	1.0	1.0	1.0	0.9	
3	3	M	7.43	0.8	0.1	0.3	0.0	
3	3	H	15.8	0.5	0.0	0.1	0.0	

Figure 1

Estimated β-errors for the Lomnicki test statistic (O) and the Cramèr-von Mises test statistic (*) for a sample size of 50 observations for different kurtosis values

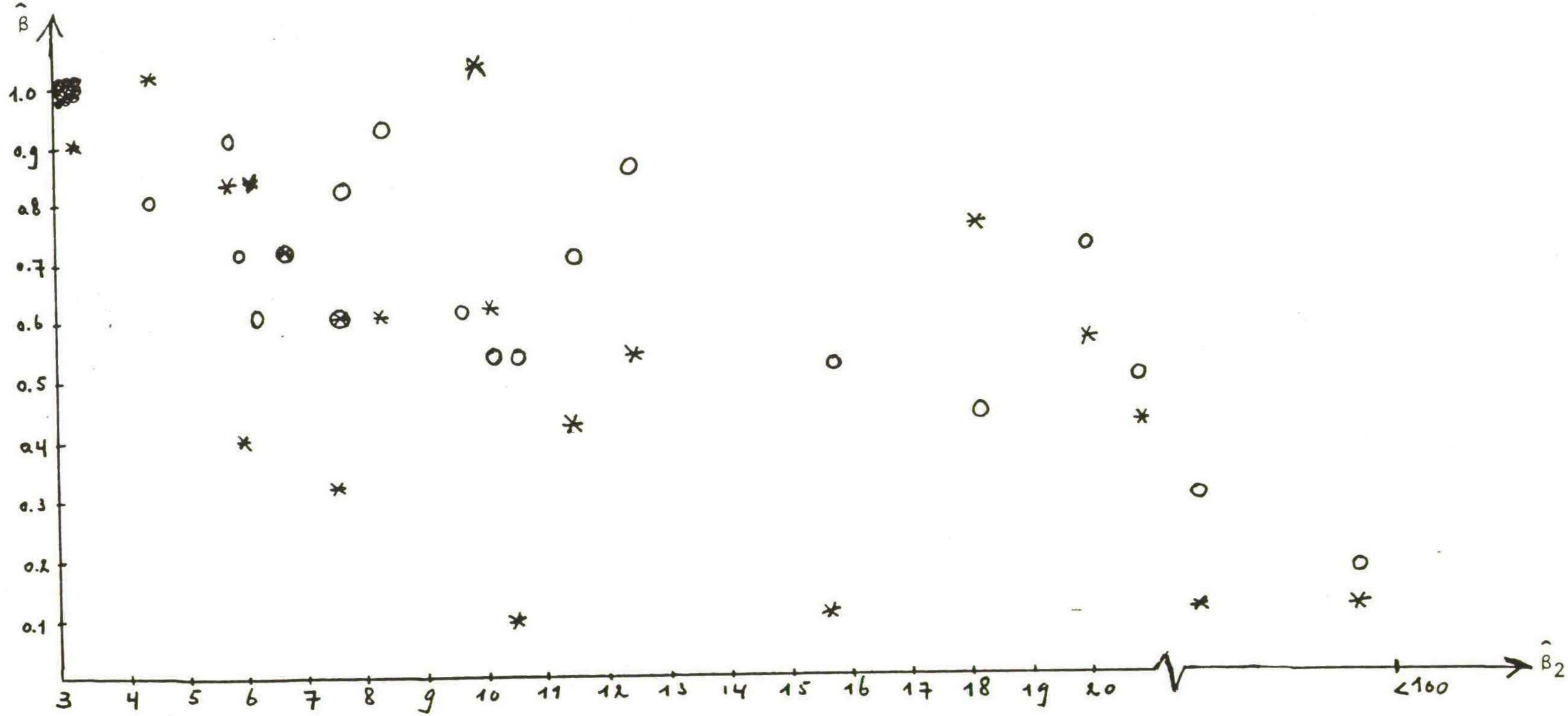
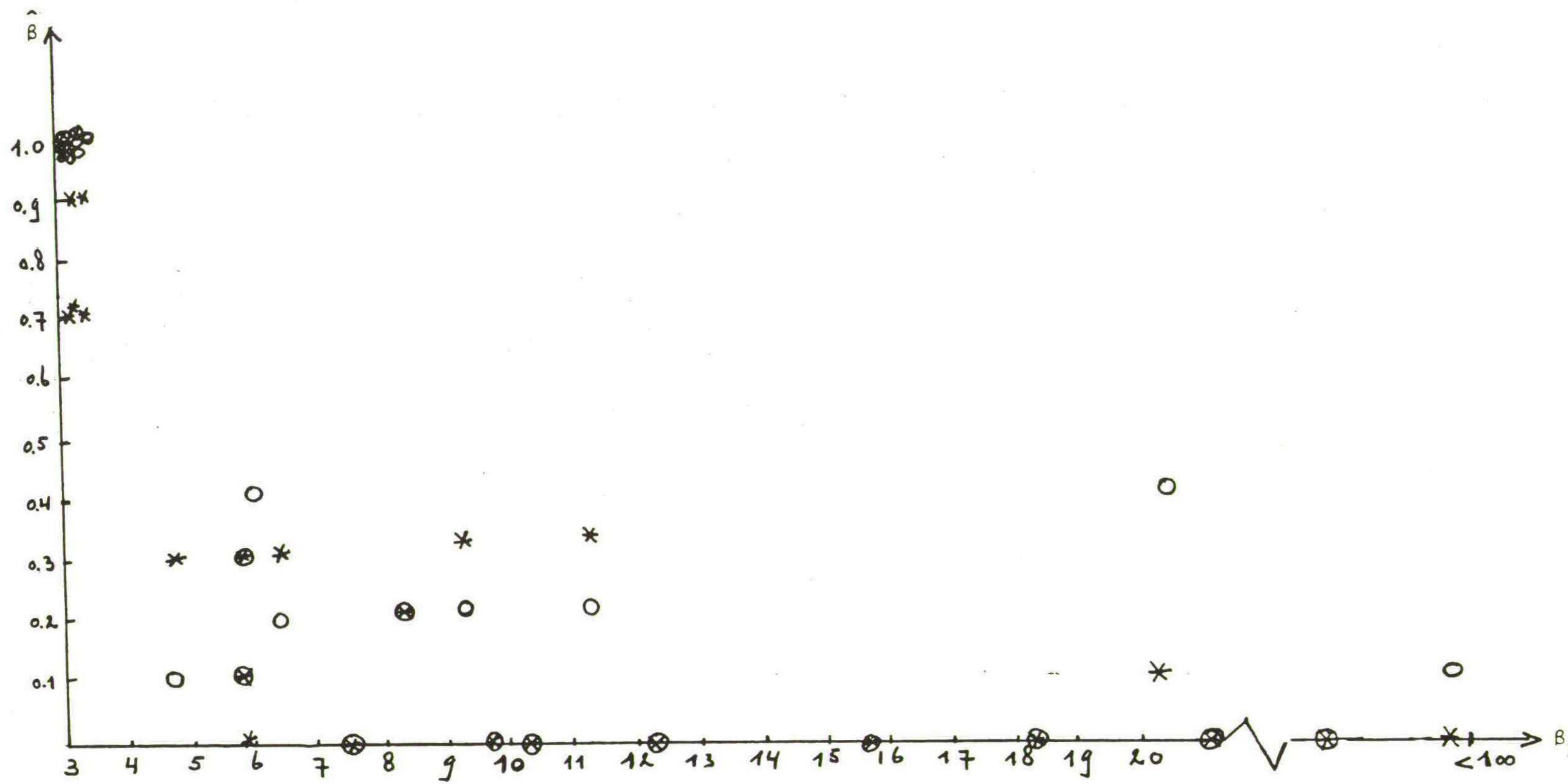


Figure 2

Estimated β-errors for the Lomnicki test statistic (O) and the Cramèr-von Mises test statistic (*) for a sample size of 200 observations for different kurtosis values



To investigate the difference in power of the two test statistics on normality we also applied the conditional sign test [4]. The sample consists of differences between the estimated β-errors for the test statistic of Lomnicki and Cramèr-von Mises. The observations will fall in one of the following classes:

+ : $\hat{\beta}$ of Cramèr-von Mises < $\hat{\beta}$ of Lomnicki

− : $\hat{\beta}$ of Cramèr-von Mises > $\hat{\beta}$ of Lomnicki

0 : $\hat{\beta}$ of Cramèr-von Mises = $\hat{\beta}$ of Lomnicki

The corresponding true probabilities are p, q and 1-p-q resp. and we test the hypothesis $H_0 : p = q$ against $H_1 : p \neq q$ (two-sided test) with level $\alpha = 0.05$. For sample sizes 50 and 200 the null hypothesis is not rejected at the above significance level.

The following alternative was also investigated: apply the conditional sign test only to those cases where the kurtosis value is larger than 4.5. Again the null hypothesis was not rejected for both the sample sizes at a significance level of 5%.

To get some impression about the possible relation between the estimated β-error and the kurtosis, we fitted a linear and an exponential regression equation to the data, which where of the following type:

$$\text{(a) } y_t = a_0 + a_1 x_t + \varepsilon_t, \quad \text{(b) } y_t = b_0 + b_1 \exp\{b_2 x_t\} + \varepsilon_t,$$

where y_t is the estimated β-error and x_t the kurtosis. The complete results will not be given here, but are available from the authors.

A summary of the results is given below:

a) Both equations lead to a significant negative relation between the kurtosis and the estimated β-error, for both tests and both sample sizes.

b) The exponential regression equation clearly has a better fit for both tests and a sample size of 200 (in terms of a much lower residual variance).

These results agree with the expectation that a high kurtosis means a large deviation from normality and so leads to a high power of the test statistics.

b. numerical problems when estimating the transformation parameters $\xi,\gamma,\lambda,\delta$

Table 6 shows the number of times that the estimation procedure of the transformation parameters was successful for specific "case". A case means a certain choice of scheme, parameter combination, size of variance and sample size. The size of the variance is denoted by L = low, M = medium, H = high (see Table 3). The number of successful estimations is at most 10, the number of replications. It is striking that most numerical problems arise for series with low variances and hence low kurtosis values (see Table 3). Fortunately, those are the cases where the Johnson S_U transformations are not interesting because the probability distributions are close to the normal distribution. Further the number of successful parameter estimations is somewhat disappointing for parameter combination 3, even for medium and large variances (and also for the same type of kurtosis values). So it is interesting to look at the position of the generated probability distributions in the (β_1,β_2)- plane (see Figure 3). Johnson [10] has shown that the S_U system generates probability distributions which are unbounded at both sides. In the (β_1,β_2)- plane the S_U system is traced by the curve of the S_L system, existing of probability distributions which are bounded at one end. Probably, parameter combination 3 exists of probability distributions, where nearly all outliers lie at one side. The other side of the S_L curve exists of probability distributions, which are bounded to both sides, and is called the S_B system.

Figure 3

The position of the generated probability distributions in the (β_1, β_2) space

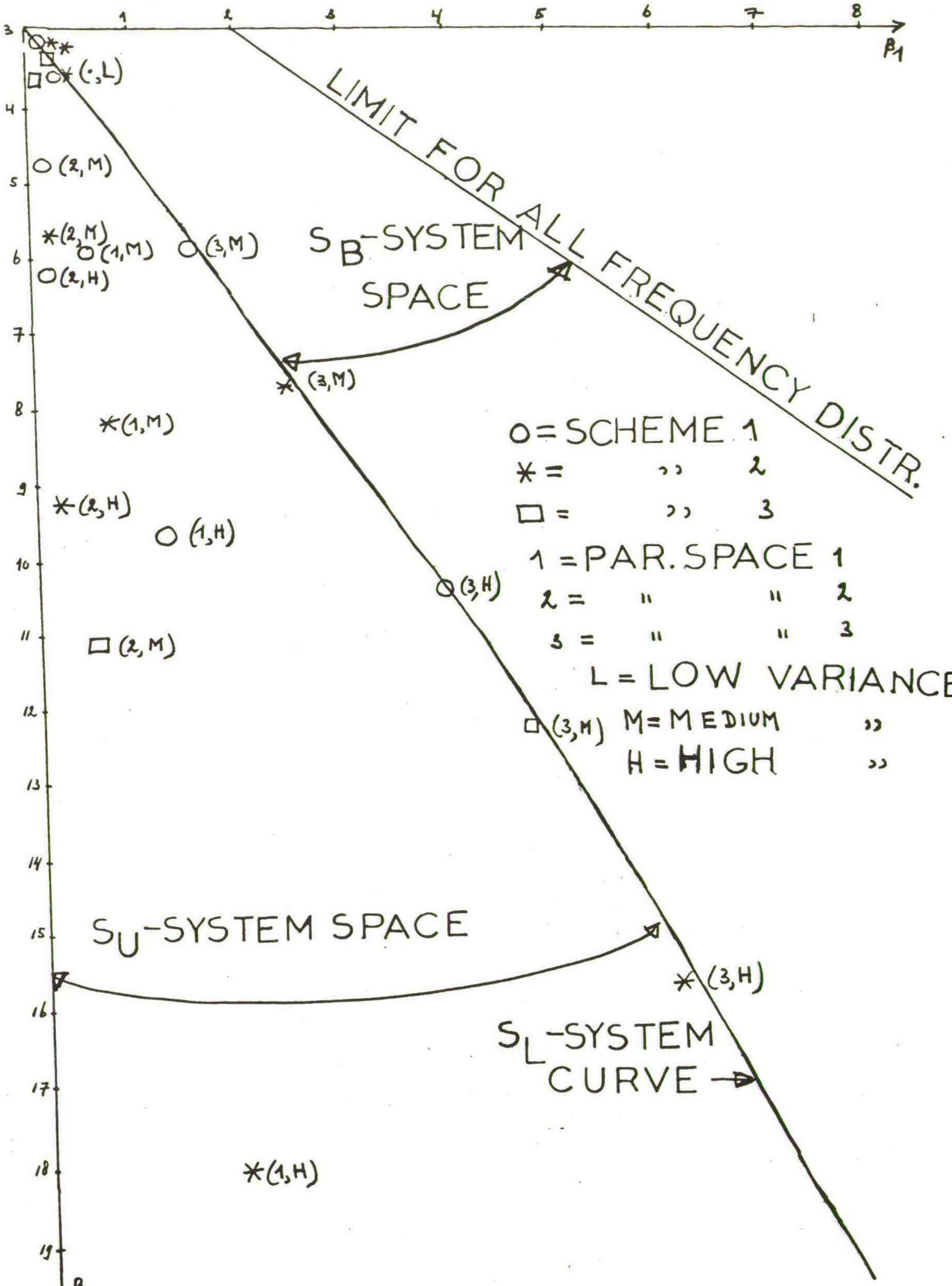


Table 6

Number of successful parameter estimations for the Johnson S_U transformation

variance		L	M	H	L	M	H	L	M	H		
scheme 1		3	6	8	2	7	7	1	7	6	50	sample size
		3	10	8	3	9	8	4	7	3	200	
scheme 2		3	7	10	2	9	7	4	6	6	50	sample size
		3	9	9	3	9	8	6	7	9	200	
scheme 3		2	7	9	1	9	8	2	6	7	50	sample size
		4	10	7	4	9	9	4	8	6	200	
		parameter combination 1			parameter combination 2			parameter combination 3				

There are some missing points in Figure 3 which could not be shown because of their extreme values.

c. Box-Jenkins structural parameters estimates per scheme

Tables 7, 8 and 9 display results for the Box-Jenkins structural parameter estimates. Per case (certain combination of scheme, parameter combination, variance and sample size) we have calculated some characteristic numbers. These numbers are:

per parameter

(1) The average relative absolute error:

$$r_{k,y} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{\alpha}_{k,y,i} - \alpha_{k,y,i}}{\alpha_{k,y,i}} \right|$$

$$r_{k,x} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{\alpha}_{k,x,i} - \alpha_{k,x,i}}{\alpha_{k,x,i}} \right|$$

where α refers to the parameter, y to the original non-normal data, x to the transformed data, k to the numbering of the parameters, and n to the number of replications.

also per parameter

(2) The differences:

$$r_{k,y} - r_{k,x} \equiv v_k.$$

When $v_k > 0$ this means that the Johnson S_U transformation has lead to an improvement of the k-th structural parameter estimate, when $v_k < 0$ the opposite holds and when $v_k = 0$ it does not matter whether the S_U transformation is applied or not.

(3) The number of successfully applied Johnson S_U transformations.

(4) The kurtosis values of the y_t series.

Table 7

Some important figures concerning the improvement of the structural parameter estimates in using the Johnson S_U transformations

(scheme 1)

SCHEME	PARAMETER COMB.	VARIANCE	SAMPLE SIZE: 50	SAMPLE SIZE: 200	coefficient 1 = 0.5			coefficient 2 = 0.2			coefficient 3 = 0.7			NUMBER OF SUCCESSFUL ESTIMATIONS	KURTOSIS
					$r_{1,y}$	$r_{1,x}$	v_1	$r_{2,y}$	$r_{2,x}$	v_2	$r_{3,y}$	$r_{3,x}$	v_3		
1	1	L	+		0.172	0.216	-0.044	2.695	2.80	-0.105	0.281	0.267	0.014	3	3.03
1	1	L		+	0.178	0.134	0.044	1.875	1.535	0.340	0.131	0.133	-0.002	3	3.03
1	1	M	+		0.580	0.290	0.290	4.195	3.425	0.770	0.313	0.309	0.004	6	20.1
1	1	M		+	0.348	0.378	-0.030	1.530	1.655	-0.125	0.130	0.164	-0.034	10	20.1
1	1	H	+		0.598	0.626	-0.028	3.430	2.995	0.435	0.186	0.288	-0.002	8	96.7
1	1	H		+	0.242	0.324	-0.082	1.725	1.615	0.110	0.186	0.213	-0.027	8	96.7

continuation of Table 7

1	2	L	+		0.250	0.296	-0.046	2.50	2.755	-0.255	0.293	0.271	0.023	2	3.01
1	2	L		+	0.170	0.134	0.036	1.825	1.530	0.295	0.133	0.133	0.0	3	3.01
1	2	M	+		0.630	0.632	-0.002	3.055	2.660	0.395	0.231	0.249	-0.018	7	11.1
1	2	M		+	0.242	0.246	-0.004	1.50	1.495	0.005	0.116	0.079	0.037	9	11.1
1	2	H	+		0.292	0.30	-0.008	3.210	2.875	0.335	0.183	0.250	-0.067	7	29.5
1	2	H		+	0.390	0.234	0.156	1.295	0.865	0.430	0.179	0.140	0.039	8	29.5
1	3	L	+		0.290	0.196	0.094	0.350	0.490	-0.140	0.316	0.313	0.003	1	3.10
1	3	L		+	0.406	0.402	0.004	1.695	1.755	-0.060	0.093	0.093	0.0	4	3.10
1	3	M	+		0.250	0.258	-0.008	4.045	2.353	1.692	0.244	0.234	0.010	7	12.4
1	3	M		+	0.296	0.246	0.050	1.035	1.575	-0.540	0.101	0.088	0.013	7	12.4
1	3	H	+		0.472	0.446	0.026	1.895	2.82	-0.925	0.270	0.283	-0.013	6	41.1
1	3	H		+	0.328	0.384	-0.056	3.250	2.510	0.740	0.340	0.163	0.177	3	41.1

Table 8

Some important figures concerning the improvement of the structural parameter estimates in using the Johnson S_U transformations

(scheme 2)

SCHEME	PARAMETER COMB.	VARIANCE	SAMPLE SIZE: 50	SAMPLE SIZE: 200	coefficient 1 = -0.7			coefficient 2 = -0.4			coefficient 3 = -0.25			coefficient 4 = -0.12			NUMBER OF SUCCESSFULL ESTIMATIONS	KURTOSIS
					$r_{1,y}$	$r_{1,x}$	v_1	$r_{2,y}$	$r_{2,x}$	v_2	$r_{3,y}$	$r_{3,x}$	v_3	$r_{4,y}$	$r_{4,x}$	v_4		
2	1	L	+		0.061	0.027	0.034	0.923	1.033	-0.11	1.952	2.420	-0.47	7.467	8.06	-0.59	3	3.01
2	1	L		+	0.073	0.057	0.016	0.313	0.323	-0.010	1.068	0.960	0.11	4.183	4.21	-0.03	3	3.01
2	1	M	+		0.30	0.343	-0.043	0.845	1.210	-0.365	2.732	2.688	0.044	11.41	11.63	-0.22	7	5.81
2	1	M		+	0.127	0.104	0.023	0.340	0.335	0.005	1.528	1.420	0.108	5.13	4.85	0.28	9	5.81
2	1	H	+		0.246	0.266	-0.020	0.980	1.035	-0.055	2.78	2.99	-0.21	11.45	12.96	-1.51	10	9.73
2	1	H		+	0.150	0.106	0.044	0.553	0.373	0.180	1.392	1.188	0.204	3.88	3.71	0.17	9	9.73
2	2	L	+		0.056	0.016	0.040	0.873	1.070	-0.197	2.920	3.48	-0.56	9.48	10.30	-0.82	2	3.01
2	2	L		+	0.073	0.057	0.016	0.300	0.318	-0.018	1.056	0.96	0.096	4.23	4.23	0.0	3	3.01
2	2	M	+		0.241	0.270	-0.029	1.113	1.103	0.010	3.104	3.30	-0.196	12.85	13.80	-0.95	9	4.68
2	2	M		+	0.097	0.091	0.006	0.395	0.313	0.082	1.080	1.03	0.05	3.12	3.54	-0.42	9	4.68

continuation of Table 8

2	2	H	+		0.226	0.307	-0.081	1.068	1.073	-0.005	1.456	1.64	-0.18	6.21	8.0	-1.80	7	6.25
2	2	H		+	0.113	0.127	-0.014	0.505	0.600	-0.095	0.896	1.45	-0.554	3.97	4.93	-0.96	8	6.25
2	3	L	+		0.201	0.200	-0.001	0.753	0.680	0.073	2.384	2.30	0.054	10.23	9.96	0.27	4	3.04
2	3	L		+	0.086	0.106	-0.020	0.635	0.653	-0.018	1.460	1.30	0.16	2.58	2.03	0.55	6	3.04
2	3	M	+		0.340	0.323	0.017	0.885	1.023	-0.138	1.908	2.64	-0.73	13.18	12.66	0.52	6	5.88
2	3	M		+	0.281	0.166	0.115	1.090	0.288	0.802	1.408	1.36	0.048	7.03	4.22	2.81	7	5.88
2	3	H	+		0.431	0.274	0.157	1.560	1.013	0.547	4.688	3.06	1.63	7.34	7.27	0.07	5	10.4
2	3	H		+	0.439	0.147	0.292	1.230	0.395	0.835	2.288	0.90	1.38	5.73	3.11	2.62	9	10.4

Table 9

Some important figures concerning the improvement of the structural parameter estimates in using the Johnson S_U transformations

(scheme 3)

SCHEME	PARAMETER COMB.	VARIANCE	SAMPLE SIZE: 50	SAMPLE SIZE: 200	coefficient 1 = 0.40			coefficient 2 = 0.90			NUMBER OF SUCCESSFUL ESTIMATIONS	KURTOSIS
					$r_{1,y}$	$r_{1,x}$	v_1	$r_{2,y}$	$r_{2,x}$	v_2		
3	1	L	+		0.165	2.010	-0.045	1.111	0.108	1.003	2	3.02
3	1	L		+	0.378	0.373	0.005	0.013	0.012	0.001	4	3.02
3	1	M	+		0.705	0.655	0.050	0.089	0.089	0.000	7	8.08
3	1	M		+	0.540	0.308	0.232	0.131	0.019	0.112	10	8.08
3	1	H	+		0.663	0.915	-0.252	0.090	0.289	-0.199	9	18.1
3	1	H		+	0.570	0.403	0.167	0.236	0.044	0.192	7	18.1
3	2	L	+		0.030	0.115	-0.085	0.127	0.126	0.001	1	3.01
3	2	L		+	0.380	0.378	0.002	0.0133	0.012	0.001	4	3.01

continuation of Table 9

3	2	M	+		0.763	0.938	-0.175	0.087	0.088	-0.001	8	5.86
3	2	M		+	0.500	0.473	0.027	0.028	0.024	0.004	9	5.86
3	2	H	+		0.503	0.485	0.018	0.122	0.114	0.008	7	9.31
3	2	H		+	0.390	0.290	0.100	0.080	0.048	0.032	9	9.31
3	3	L	+		0.290	0.445	-0.155	0.120	0.127	-0.007	1	3.06
3	3	L		+	0.705	0.763	-0.058	0.037	0.021	0.016	4	3.06
3	3	M	+		0.770	0.850	-0.080	0.117	0.091	0.026	5	7.43
3	3	M		+	0.650	0.393	0.257	0.382	0.037	0.345	8	7.43
3	3	H	+		0.515	0.403	0.112	0.103	0.099	0.004	6	15.8
3	3	H		+	1.080	0.523	0.492	0.470	0.098	0.372	6	15.8

Remarks by tables 7, 8 and 9:

The results in these tables are restricted to successfully applied Johnson S_U transformations. A global impression from the tables is, that when the data are in fact non-normally distributed, the Johnson S_U transformation does in general not lead to significantly improved structural parameter estimates. When the information in the tables is studied separately for small and large sample sizes and for different kurtosis values, the conclusions are as follows: enlarging the sample size or the kurtosis values, generally improves the structural parameter estimates via the Johnson S_U transformation only for scheme 2 and 3, for scheme 1 it remains nearly the same. Other tables which will not be given here indicate that the Box-Jenkins estimation procedure for the structural parameters is fairly robust against non-normality.

d. The possible relation between the quality of the transformation parameter estimates and that of the structural parameter estimates

An interesting question is if an improvement on the structural parameter estimates is due to the good quality of the transformation parameter estimates.

To answer this question, we first concider relative and absolute deviations of the parameter estimates from their true values, before and after transformation, and then determined how often an overall improvement or worsening of the structural parameter estimates occurred after transformation. This has lead to the following global impression: We could not find any clear relation between the quality of the structural parameter estimates after Johnson S_U transformation and the quality of the transformation parameter estimates.

7. CONCLUSIONS

When testing normality of correlated data that are in fact sampled from Johnson S_U distributions, the test statistics of Lomnicki and Cramèr-von Mises have a comparable power function, at least for the cases which we investigated. For both statistics the power is not good for a sample size of 50 observations, but for a sample size of 200 observations and large kurtosis it is reasonable. Further it should be noted that numerical problems arise in the estimation of the Johnson S_U transformation parameters, when the generated probability distributions lie near the normal one or near the boundary of the Johnson S_L distributions. Finally we found the following interesting result: When the data are sampled from a Johnson S_U distribution, the S_U transformed data do in general not lead to improved structural parameter estimates in ARMA-models, suggesting that the Box-Jenkins structural parameter estimates are fairly robust against the type of non-normality which we studied. This impression is strengthened by the fact that we could not find a relation between the quality of the structural parameter estimates after Johnson S_U transformation and the quality of the transformation parameter estimates.

APPENDIX A - FORMULAS FOR THE FIRST FOUR MOMENTS OF THE INVERSE JOHNSON S_U TRANSFORM

Assume $y = \sinh\left(\frac{z-\gamma}{\delta}\right)$, with $z \in N(0,1)$.

For the first central moments of y we have [10]:

$$\mu_1 = -\sqrt{\omega}\,\sinh\Omega$$

$$\mu_2 = \frac{1}{2}(\omega-1)\,(\omega\cosh(2\Omega)+1)$$

$$\mu_3 = -\frac{1}{4}(\omega-1)^2 \sqrt{\omega}\,\{\omega(\omega+2)\sinh(3\Omega) + 3\sinh\Omega\}$$

$$\mu_4 = (\omega-1)^2\,\{d_4\cosh(4\Omega) + d_2\cosh(2\Omega) + d_0\}$$

with $\omega = \exp(1/\delta^2)$, $\Omega = \gamma/\delta$, $d_4 = \frac{1}{8}\,\omega^2(\omega^4+2\omega^3+3\omega^2-3)$,
$d_2 = \frac{1}{2}\,\omega^2(\omega+2)$, $d_0 = 3(2\omega+1)/8$.

Now assume: $y = \xi + \lambda\sinh(\frac{(z-\gamma)}{\delta})$, with $z \in N(0,1)$, then the first four central moments of y are:

$$\mu_1^* = \lambda\,\mu_1 + \xi,\ \mu_2^* = \lambda^2\,\mu_2,\ \mu_3^* = \lambda^3\,\mu_3,\ \mu_4^* = \lambda^4\,\mu_4.$$

At last, assume $y = \xi + \lambda\sinh(\frac{(x-\gamma)}{\delta})$, with $x \in N(0,\sigma^2)$, then:

$$z = x/\sigma \in N(0,1) \text{ and } y = \xi + \lambda\sinh(\frac{z-\gamma/\sigma}{\delta/\sigma}).$$

The central moments of y are now: μ_1^*, μ_2^*, μ_3^*, μ_4^*, where in this case:

$$\omega = \exp(\frac{\sigma^2}{\delta^2}) \text{ and } \Omega = (\frac{\gamma}{\delta}).$$

APPENDIX B - GENERATING NON-NORMAL SEQUENCES FROM ARMA SCHEMES VIA NON-NORMAL NOISE PROCESSES

Let us assume that x_t satisfies an infinite moving average scheme or can be written in this form: $x_t = \sum_{j=0}^{\infty} c_j\,\varepsilon_{t-j}$, where $\{\varepsilon_t\}$ is a white noise process, with $E\{\varepsilon_t\} = 0$. Then it can be shown that the skewness and the kurtosis of x_t can be expressed in the equivalent expressions of the $\{\varepsilon_t\}$ process:

$$\sqrt{\beta_1(x)} = \left(\sum_{i=0}^{\infty} c_i^3\right)\left(\sum_{i=0}^{\infty} c_i^2\right)^{-\frac{3}{2}} \sqrt{\beta_1(\varepsilon)}$$

$$\beta_2(x) = \left(\sum_{i=0}^{\infty} c_i^4\right)\left(\sum_{i=0}^{\infty} c_i^2\right)^{-2} \beta_2(\varepsilon) + 6\left(\sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty} c_i^2 c_j^2\right) \times$$

$$\left(\sum_{i=0}^{\infty} c_i^2\right)^2, \text{ where } \sqrt{\beta_1(x)} := (\mu_3)(\mu_2)^{-\frac{3}{2}} = (\mu_3')(\mu_2')^{-\frac{3}{2}}$$

and $\beta_2(x) := (\mu_4)(\mu_2)^{-2} = (\mu_4')(\mu_2')^{-2}$, as $E\{x_t\} = 0$.

In case of normal distributed noise we have $\sqrt{\beta_1(\varepsilon)} = 0$ and $\beta_2(\varepsilon) = 3$. When $\{\varepsilon_t\}$ is a non-normal process, then we will investigate the skewness and kurtosis measures for the $\{x_t\}$ process, assuming the $\{x_t\}$ satisfy the ARMA-schemes in section 2. Therefore we have to rewrite the ARMA-schemes in the infinite moving average form and then use the above formulas.

Table B-1
Values of $\sqrt{\beta_1(x)}$ and $\beta_2(x)$ for the ARMA-schemes in section 2, by assuming different values for $\sqrt{\beta_1(\varepsilon)}$ and $\beta_2(\varepsilon)$.

Table of $\sqrt{\beta_1(x)}$			
$\sqrt{\beta_1(\varepsilon)}$	Model 1	Model 2	Model 3
0.5	0.155	0.178	0.042
1	0.310	0.356	0.083
2	0.620	0.712	0.167
3.3	1.024	1.174	0.275
5	1.551	1.780	0.417
6	1.861	2.135	0.500

Table of $\beta_2(x)$

$\beta_2(\varepsilon)$	Model 1	Model 2	Model 3
3	3	3	3
5	3.249	4.093	3.771
7	3.499	5.187	4.541
9	3.748	6.281	5.311
11	3.998	7.374	6.082
13	4.247	8.468	6.852
20	5.121	12.297	9.550
40	7.742	23.782	17.641

From the above Table B-1 we can see that the above generating mechanism for non-normal sequences cannot be recommended beause a strongly non-normal $\{\varepsilon_t\}$ process with high $(\sqrt{\beta_1},\beta_2)$-values is reduced to an $\{x_t\}$ process with much lower $(\sqrt{\beta_1},\beta_2)$-values, probably under influence of the central limit theorem.

REFERENCES

[1] G.E.P. Box and G.M. Jenkins, *Time Series Analysis, Forecasting and Control*, 2nd ed. (Holden-Day, San Francisco, 1976).

[2] R.P. Brent, A Gaussian pseudo-random number generator, *Communications of the A.C.M.* 17 (1974) 704-706.

[3] E.F. Fama, The behavior of stock market prices, *J. of Business 38* (1965) 34-105.

[4] J. Hájek, *A Course of Non-parametric Statistics* (Holden-Day, San Francisco, 1969).

[5] R.M.J. Heuts, The use of non-linear transformations in ARIMA-models when the data are non-Gaussian distributed. Reeks "Ter Discussie", No. 82.08, Dept. of Econometrics, Tilburg University, Tilburg (1982).

[6] I.D. Hill, Normal-Johnson and Johnson-normal transformations, *Applied Statistics* 25 (1976) 190-192.

[7] I.D. Hill, R. Hill and R.L. Holder, Fitting Johnson curves by moments, *Applied Statistics* 25 (1976) 180-189.

[8] I.D. Hill and R.E. Wheeler, Fitting Johnson curves by moments, *Applied Statistics* 30 (1981) 106.

[9] G.M. Jenkins and Partners, *Jenasys: Box-Jenkins Forecasting Packages* (Gwilym Jenkins and Partners, Lancaster, 1981).

[10] N.L. Johnson, Systems of frequency curves generated by methods of translation, *Biometrika* 36 (1949) 149-176.

[11] N.L. Johnson, Tables to facilitate fitting S_U frequency curves, *Biometrika* 52 (1965) 547-558.

[12] J.F. Lawless, *Statistical Models and Methods for Lifetime Data* (Wiley, New York, 1982).

[13] D.C.M. Leslie, Determination of parameters in the Johnson system of probability distributions, *Biometrika* 46 (1959) 229-231.

[14] Z.A. Lomnicki, Test for departure from normality in the case of linear stochastic processes, *Metrika* 4 (1961) 37-62.

[15] B. Mandelbrot, The variation of certain speculative prices, *J. of Business* 36 (1963) 394-419.

[16] M.A. Stephens, EDF Statistics for goodness of fit and some comparisons, *J. Amer. Statist. Assoc.* 69 (1974) 730-737.

IN 1982 REEDS VERSCHENEN:

107 Aart J. de Zeeuw
Hierarchical decentralized optimal control in econometric policy models.

108 Arie Kapteyn en Tom Wansbeek
Identification in Factor Analysis.

109 G. van der Laan en A.J.J. Talman
Simplical Algorithms for finding Stationary Points, a unifying description

110 Pieter Boot
Ekonomiese betrekkingen tussen Oost en West Europa.

111 B.B. van der Genugten
The asymptotic behaviour of the estimated generalized least squares method in the linear regression model.

112 Jack P.C. Kleijnen / Anton J. van Reeken
Principles of computer charging in a university-like organization.

113 H. Tigelaar.
The informative sample size for dynamic multiple equation systems with moving average errors.

114 Drs. H.G. van Gemert, Dr. R.J. de Groof en Ir. A.J. Markink
Sektorstruktuur en Economische Ontwikkeling

115 P.H.M. Ruys
The tripolar model: a unifying approach to change.

116 Aart J. de Zeeuw
Policy solutions for Mini-Interplay, a Linked Model for two Common Market Countries.

IN 1982 REEDS VERSCHENEN (vervolg):

117 F.J.M. van Doorne en P.H.M. Ruys
Die Struktur einer Sprechhandlung in Habermas' Forschungsprogramm.
Formale Analyse mit den Mitteln des tripolaren Modells.

118 A.J.J. Talman en G. van der Laan
Simplicial Approximation of Solutions to the Non-Linear Complementarity Problem

119 Pieter Boot
The East German Planning System Reconsidered

120 J.P.C. Kleijnen
Regression Metamodel Summarization of Model Behaviour

121 P.G.H. Mulder en A.L. Hempenius
Evaluating life time in a competing risks model for a chronic desease.

122 A.J.J. Talman en G. van der Laan
From Fixed Point to Equilibrium

123 J.P.C. Kleijnen
Design of simulation experiments.

124 H.L. Theuns en A.M.L. Passier-Grootjans
Internationaal toerisme; een gids in de algemene basisliteratuur en het bronnenmateriaal.

125 J.H.F. Schilderinck
Interregional Structure of the EUROPEAN community
Part I: Imports and Exports, Sub-divided by Countries Aggregated According the Branches of the European Community Interregional Input-Outputtables 1959, 1965, 1970 and 1975.

IN 1983 REEDS VERSCHENEN:

126 Harry H. Tigelaar
Identification of noisy linear systems with multiple arma inputs.

127 J.P.C. Kleijnen
Statistical Analysis of Steady-State Simulations: Survey of Recent Progress.

128 Aart J. de Zeeuw
Two notes on Nash and Information.

129 H.L. Theuns en A.M.L. Passier-Grootjans
Toeristische ontwikkeling - voorwaarden en systematiek; een selectief literatuuroverzicht.

130 Joseph Plasmans en Valère Somers
A Maximum Likelihood Estimation Method of a Three Market Disequilibrium Model.